

THAT WHICH IS CLAIMED IS:

1. A method of controlling a cache of distributed data, comprising:
dynamically determining whether and/or where to cache the distributed
data based on characteristics of the data, characteristics of the source of the data
5 and characteristics of the cache so as to provide an indication of whether to cache
the data; and
selectively caching the data based on the indication.
2. The method of Claim 1, wherein the characteristics of the data
10 comprise how often the data is accessed.
3. The method of Claim 1, wherein the characteristics of the source of
the data comprise how long it takes to recompute the data and/or how long it takes
to replicate the data.
15
4. The method of Claim 1, wherein the characteristics of the cache
comprise how long it takes to retrieve a cached item.
5. The method of Claim 1, wherein dynamically determining whether
20 and/or where to cache the distributed data, comprises:
determining a predicted maximum number of cache accesses;
determining a predicted maximum time consumed by processing cache hits
corresponding to a cache entry corresponding to the distributed data;
determining a time (r) to replicate the distributed data;
25 determining time (c) to generate the distributed data; and
setting the indication to indicate caching the distributed data if the sum of
the time to generate the distributed data, the time to replicate the distributed data
and the predicted maximum time consumed by processing cache hits is less than
the product of the predicted maximum number of cache accesses and the time to
30 generate the distributed data.

6. The method of Claim 1, wherein setting the indication is repeatedly performed for a time (r) that is equal to a time to retrieve the distributed data from a local cache, a time to replicate the distributed data in a cluster, and a time to offload the distributed data to disk, to thereby determine whether and where to
5 cache the distributed data.

7. The method of Claim 5, further comprising:
determining a time to live (TTL) for the cache entry corresponding to the distributed data;
10 determining a time (h) to process a cache hit corresponding to the distributed data;
determining a predicted frequency (f) of cache accesses for the cache entry corresponding to the distributed data;
wherein determining a predicted maximum number of cache access
15 comprises determining TTL*f; and
wherein determining a predicted maximum time consumed by processing cache hits corresponding to a cache entry corresponding to the distributed data comprises determining $h*(TTL*f)-1$.

20 8. The method of Claim 1, wherein the cache comprises a disk cache and wherein caching the data comprises offloading cached memory contents to the disk cache.

9. The method of Claim 5, wherein determining a predicted maximum
25 number of cache access comprises monitoring cache accesses to determine an update rate of cache entries corresponding to the distributed data.

10. The method of Claim 7, wherein determining a time (h) to process a cache hit corresponding to the distributed data comprises monitoring cache
30 accesses to determine the time (h).

11. The method of Claim 5, wherein determining a time (r) to replicate the distributed data comprises monitoring data replication operations to determine the time (r).

5 12. The method of Claim 5, wherein determining time (c) to generate the distributed data comprises monitoring generation of the distributed data to determine the time (c).

10 13. A system for controlling a cache of distributed data, comprising:
means for dynamically determining whether and/or where to cache the distributed data based on characteristics of the data, characteristics of the source of the data and characteristics of the cache so as to provide an indication of whether to cache the data; and
means for selectively caching the data based on the indication.

15 14. The system of Claim 13, wherein the means for dynamically determining whether and/or where to cache the distributed data, comprises:
means for determining a predicted maximum number of cache accesses;
means for determining a predicted maximum time consumed by processing cache hits corresponding to a cache entry corresponding to the distributed data;
means for determining a time (r) to replicate the distributed data;
means for determining time (c) to generate the distributed data; and
means for setting the indication to indicate caching the distributed data if the sum of the time to generate the distributed data, the time to replicate the distributed data and the predicted maximum time consumed by processing cache hits is less than the product of the predicted maximum number of cache accesses and the time to generate the distributed data.

20 15. The system of Claim 14, further comprising:
means for determining a time to live (TTL) for the cache entry corresponding to the distributed data;

means for determining a time (h) to process a cache hit corresponding to the distributed data;

means for determining a predicted frequency (f) of cache accesses for the cache entry corresponding to the distributed data;

5 wherein the means for determining a predicted maximum number of cache access comprises means for determining $\text{TTL} \cdot f$; and

wherein the means for determining a predicted maximum time consumed by processing cache hits corresponding to a cache entry corresponding to the distributed data comprises means for determining $h \cdot (\text{TTL} \cdot f) - 1$.

10

16. The system of Claim 13, wherein the cache comprises a disk cache and wherein the means for selectively caching the data comprises means for offloading cached memory contents to the disk cache.

15

17. A computer program product for controlling a cache of distributed data, comprising:

a computer readable medium having computer readable program code embodied therein, the computer readable program code comprising:

20 computer readable program code configured to dynamically determine whether and/or where to cache the distributed data based on characteristics of the data, characteristics of the source of the data and characteristics of the cache so as to provide an indication of whether to cache the data; and

computer readable program code configured to selectively cache the data based on the indication.

25

18. The computer program product of Claim 17, wherein the computer readable program code configured to dynamically determine whether and/or where to cache the distributed data, comprises:

30 computer readable program code configured to determine a predicted maximum number of cache accesses;

computer readable program code configured to determine a predicted maximum time consumed by processing cache hits corresponding to a cache entry corresponding to the distributed data;

5 computer readable program code configured to determine a time (r) to replicate the distributed data;

computer readable program code configured to determine time (c) to generate the distributed data; and

10 computer readable program code configured to set the indication to indicate caching the distributed data if the sum of the time to generate the distributed data, the time to replicate the distributed data and the predicted maximum time consumed by processing cache hits is less than the product of the predicted maximum number of cache accesses and the time to generate the distributed data.

19. The computer program product of Claim 18, further comprising:
15 computer readable program code configured to determine a time to live (TTL) for the cache entry corresponding to the distributed data;

computer readable program code configured to determine a time (h) to process a cache hit corresponding to the distributed data;

20 computer readable program code configured to determine a predicted frequency (f) of cache accesses for the cache entry corresponding to the distributed data;

wherein the computer readable program code configured to determine a predicted maximum number of cache access comprises computer readable program code configured to determine $TTL \cdot f$; and

25 wherein the computer readable program code configured to determine a predicted maximum time consumed by processing cache hits corresponding to a cache entry corresponding to the distributed data comprises computer readable program code configured to determine $h \cdot (TTL \cdot f) - 1$.

30 20. The computer program product of Claim 17, wherein the cache comprises a disk cache and wherein the computer readable program code

configured to selectively cache the data comprises computer readable program code configured to offload cached memory contents to the disk cache.